Nucleophilic catalysis of acylhydrazone equilibration for protein-directed dynamic covalent chemistry

Published in final edited form as:
Nat Chem. 2010 June ; 2(6): 490–497. doi:10.1038/nchem.658.

Nucleophilic catalysis of acylhydrazone equilibration for protein-directed dynamic covalent chemistry

Venugopal T. Bhat[1,†], **Anne M. Caniard**[1,†], **Torsten Luksch**[2], **Ruth Brenk**[2], **Dominic J. Campopiano**[1,*], and **Michael F. Greaney**[1,*]

[1] EastChem, School of Chemistry, University of Edinburgh, King's Buildings, West Mains Road, Edinburgh EH9 3JJ, UK

[2]College of Life Sciences, University of Dundee, James Black Centre, Dow Street, Dundee DD1 5EH, UK

Abstract

Dynamic covalent chemistry uses reversible chemical reactions to set up an equilibrating network of molecules at thermodynamic equilibrium, which can adjust its composition in response to any agent capable of altering the free energy of the system. When the target is a biological macromolecule, such as a protein, the process corresponds to the protein directing the synthesis of its own best ligand. Here, we demonstrate that reversible acylhydrazone formation is an effective chemistry for biological dynamic combinatorial library formation. In the presence of aniline as a nucleophilic catalyst, dynamic combinatorial libraries equilibrate rapidly at pH 6.2, are fully reversible, and may be switched on or off by means of a change in pH. We have interfaced these hydrazone dynamic combinatorial libraries with two isozymes from the glutathione S-transferase class of enzyme, and observed divergent amplification effects, where each protein selects the best-fitting hydrazone for the hydrophobic region of its active site.

Dynamic covalent chemistry (DCC) uses reversible chemical reactions to set up equilibrating assemblies of molecules at thermodynamic equilibrium[1-4]. The resultant dynamic combinatorial library (DCL) is responsive to the addition of a template, which will selectively amplify the best binding compounds from the equilibrium distribution. The essence of the concept lies in the subsequent adjustment of the DCL equilibrium, which will express more of the best binding compounds at the expense of the poorer ones. A DCL is thus adaptive and capable of evolutionary behaviour, whereby individual components are either amplified or reduced in response to template-directed binding events. These concepts have been applied to diverse problems in biological and medicinal chemistry[5-11], synthetic receptor–ligand interactions[12-16], self-replication[17-19], complex molecule synthesis[20-22] and materials science[23,24]. Taken together, they represent the best characterized examples to date of systems chemistry, which looks to synthesize complex molecular networks and study

*Correspondence and requests for materials should be addressed to D.J.C. and M.F.G. Dominic.Campopiano@ed.ac.uk; Michael.Greaney@ed.ac.uk.
†These authors contributed equally to this work.

Author contributions

V.T.B., A.M.C., D.J.C. and M.F.G. conceived and designed the experiments, V.T.B. and A.M.C. performed the experiments, and T.L., R.B. and A.M.C. carried out molecular modelling. All authors discussed the results and co-wrote the manuscript.

The authors declare no competing financial interests.

Supplementary information and chemical compound information accompany this paper at www.nature.com/naturechemistry.

Europe PMC Funders Author Manuscripts

their properties and behaviour in macrocosm, rather than as a sum of their individual components[25,26].

We are interested in DCC systems that use a biological molecule, such as a protein, to template assemblies of small molecules at dynamic equilibrium[27]. Here, the DCC experiment provides a method for discovering, studying and ranking novel protein ligands, concepts fundamental to medicinal chemistry. In these terms, the DCC process bridges the gap between targeted chemical synthesis of drug candidates and their biological binding assay, meshing the two processes into a single step in which the structure of the biological target directs the assembly of its own best inhibitor *in situ*.

A particular challenge for DCC in biological systems lies in the implementation of a suitable reversible reaction that can operate effectively under the physiological conditions required by the biotemplate. Lehn has defined two limiting cases for DCL construction: adaptive and pre-equilibrated DCC[28]. The adaptive DCL represents the ideal scenario, where the DCL chemistry is fully compatible with the biological target and the ensuing binding events control the evolution of the DCL composition. Pre-equilibrated DCL refers to the cases where the reversible chemistry used to constitute the DCL is not compatible with the biological target, meaning that the DCL and the target must be separated in some manner. This results in static libraries in which the molecular recognition events that control DCL composition are lost. Given the challenges associated with conducting fast, freely reversible chemistry under physiological conditions, it is not surprising that methods for true adaptive DCL generation are limited, with the majority of successful systems using sulfur-based transformations such as disulfide bond formation or thiol conjugate addition[29-32]. The development of new methods for adaptive DCLs is thus central to the application of DCC to biological systems, as the chemistry will define the target scope and range of available DCL components.

The reversible formation of C=N imine-type linkages emerged early on as a DCL-forming reaction[33]. The ready availability of diverse carbonyl and amine building blocks, plus the extensive precedent of imine formation in biochemical systems, makes it an ideal candidate reaction. However, the inherent instability of imines in aqueous solution presents serious analytical and isolation problems in the DCC context. The solution to this in the field of biological DCC has been to construct pseudo-adaptive DCLs where the imine linkage is reduced in situ to an amine with an external hydride source. The resulting library contains static amine components that can correspond to the imines in binding affinity, although both false-positive and false-negative results are possible. In addition, the introduction of an *in situ* reduction step complicates the DCL equilibration and makes it difficult to distinguish between genuine thermodynamic selection of the best binders and selection of those compounds that are kinetically favoured.

An advance on simple imine formation in DCC came from the Sanders group, who introduced acylhydrazones as reversible linkages[34]. The reaction has proven to be an excellent balance between facile reversibility and product stability; the acylhydrazone products formed are stable to analysis and isolation, and the reaction has very good equilibration properties, as is made evident by its application to a large number of elegant abiological DCC studies subsequently reported by the Sanders group[35-37]. It has not, however, been generally possible to apply this reaction directly to adaptive biological DCC systems because of the acidic pH required for reversibility to occur in a reasonable timeframe (pH < 4)[38,39]. A single elegant study from Poulsen has shown that slow equilibration of acylhydrazones, taking one week at pH 7.2, can be accelerated in the presence of the enzyme carbonic anhydrase, enabling *in situ* identification of binders using mass spectrometry[40]. We were keen to apply this proven reaction to our DCC studies of

enzymes, and reasoned that it could be harnessed as a powerful tool for biological investigation if a suitable catalyst could be found to accelerate the equilibration. Nucleophilic catalysis of semicarbazone formation using aniline derivatives was established in classic work from Jencks in the 1960s, and recently applied to hydrazone and oxime formation in peptide ligation systems by Dawson[41-44]. We reasoned that an additive such as aniline could promote the equilibration of acylhydrazides and aldehydes at pH values closer to the physiological window required by biological targets in DCC (Fig. 1).

We began by reacting aldehyde 1 (Fig. 2), related to the known glutathione S-transferase (GST) substrate chlorodinitrobenzene (CDNB, see below), with an excess of the ten aryl hydrazides **2a–2j** at room temperature. The hydrazides were chosen to randomly display aryl and heteroaryl groups and featured eight acyl and two sulfonyl hydrazides (**2d** and **j**). Equilibration at pH 6.2 was slow, and only two of the ten possible hydrazones could be observed by high-performance liquid chromatography (HPLC) after 1 h (Fig. 2c). Notably, there was a significant amount of free aldehyde 1 present throughout the reaction, despite the presence of excess amounts of the ten different hydrazides. Equilibrium was not complete after 48 h, and required incubation for a further 5 days until the library composition reached a steady-state composition with signals for each of the ten hydrazones **3a–3j** being clearly identified. In contrast, repeating the experiment in the presence of excess aniline produced a far higher rate of equilibration. A distribution of acylhydrazones was observed after initial mixing and HPLC sampling, and complete equilibration of the ten components was observed after just 6 h (Fig. 2d). Aldehyde **1** could not be detected following initial mixing, indicating that it was continually being sequestered as an acylhydrazone component, reflecting the faster exchange processes operating in the presence of aniline (see Supplementary Information for a study on the effect of varying aniline concentration on rates of hydrazone formation). We demonstrated the reversibility of the DCL by generating it from a different starting composition, hydrazone **3g** plus the nine other hydrazides and aniline. An identical equilibrium distribution to Fig. 2 was observed, indicating true thermodynamic equilibrium. A second control experiment confirmed the reversibility of the DCL through the addition of excess hydrazide **2b** to the preequilibrated DCL, which resulted in a large amplification of the corresponding acylhydrazone **3b** (see Supplementary Information).

Having established that aniline could act as an effective nucleophilic catalyst for hydrazone DCC formation at both a pH and timeframe reasonable for biomolecule stability, our next step was to introduce proteins to the DCL. Our target chosen for DCC interrogation was the GST enzyme superfamily[45]. The GSTs are responsible for cell detoxification, catalysing the conjugation of glutathione (GSH) to a wide variety of xenobiotic electrophiles, thereby protecting the cell from cytotoxic and oxidative stress. We have previously developed thiol conjugate addition DCLs directed towards GST inhibition, and successfully interfaced the enzyme with small molecules so that it controlled library evolution[27]. The GSTs are well suited to exploration using DCC methods, being well-characterized, robust proteins having nascent medicinal chemistry application[46,47]. There are relatively few ligands reported in the literature for GST binding—a plus point, as it would enable us to use DCC as a genuine discovery tool for new binding motifs, rather than as a proof-of-principle process for confirming the binding ability of known ligands. The cytoplasmic GSTs are inherent dimers with active sites composed of residues from both monomers, bifurcating between a highly conserved G-site, which binds the endogenous ligand GSH, and an H-site, which binds hydrophobic substrates for GSH conjugation (Fig. 3). This bisubstrate architecture is particularly appropriate for DCC interrogation, given that the method essentially uses a reversible linkage to couple two sets of fragment structures together[48]. Furthermore, within the GST superfamily, the large, heterogeneous H-sites are functionally evolved to

accommodate many different hydrophobic substrates for conjugation, a classically difficult architecture to investigate using orthodox structure-based drug-design methods.

We prepared two recombinant GST isozymes as targets, SjGST from the helminth worm *Schistosoma japonicum*, a drug target in tropical disease[49], and hGST P1-1, a human isoform that has been targeted in the treatment of chemotherapy drug resistance[50]. An initial control experiment with SjGST established that the enzyme retained GSH conjugation activity in the presence of aniline (up to 20 mM). The acylhydrazone DCL prepared in Fig. 2 was then interfaced with the two protein targets and amplification was measured (Fig. 4). Both DCLs demonstrated strikingly clear amplification of hydrazone components; thiophene acylhydrazone **3g** was selected by SjGST and *t*-butylphenyl hydrazone **3c** by hGST P1-1.

Synthesizing the DCL in the presence of bovine serum albumin (BSA, 1 equiv.) as a control experiment produced no measurable amplification of any component, indicating the GST enzymes as being responsible for component amplification. We further demonstrated that amplified components were bound in the target H-region of the active site of the enzyme by performing conjugation experiments with GSH. Conjugation of GSH to the aryl chloride group in hydrazones **3** by means of S_NAr substitution is a slow reaction at pH 6.2, taking several days. In the presence of catalytic amounts of SjGST, however, rapid formation of the S_NAr conjugation adduct for hydrazone **3g** was observed at pH 6.2. The amplified hydrazone can thus act as a substrate for SjGST and binds in the targeted H-site.

The amplified hydrazones were re-synthesized and assayed against both GSTs and found to be inhibitors of GSH conjugation of CDNB, but poor solubility prevented the determination of accurate IC_{50} values at the higher concentrations necessary to assay weak binding compounds. To solve this problem, and simultaneously increase the potency of our DCL components, we conjugated GSH to aldehyde **1** using an S_NAr reaction. We anticipated that the highly soluble GSH tripeptide motif would act as an 'anchor' at the G-site, enabling exploration of the H-site with assorted hydrazide fragments. This approach, in which a known enzyme–substrate interaction is used for inhibitor discovery, is well exemplified in classical medicinal chemistry drug design, GST inhibition[51] and DCC methods. The IC_{50} value for SjGST inhibition of the anchored fragment **4** was measured in the CDNB conjugation assay as 280 μM.

Initial DCC experiments using GS-conjugated aldehyde 4 and the same ten hydrazides used previously confirmed the utility of aniline as a nucleophilic catalyst (Fig. 5). Equilibration was complete in 6 h, compared to 4 days in the absence of aniline, and each of the ten acylhydrazones were clearly identified by liquid chromatography-mass spectrometry (LC-MS) (see Supplementary Information). As before, clear amplifications could be observed for both GST targets: in each case the same hydrazide fragment was selected as the best binder, thiophene (**5g**) for SjGST and *t*-butylphenyl (*5c*). Both components were amplified to over 300% of their concentrations in the blank DCL, at the expense of nearly all other competing hydrazones. Also of note, the anisyl sulfonylhydrazone **5j** underwent ~100% amplification using hGST P1-1 as the only other positively selected component. The most significant reductions in equilibrium concentrations occurred for **5b**, **f** and **i** (SjGST) and **5f**, **g** and **i** (hGST P1-1).

The GST-directed DCLs were synthesized with the protein present from the beginning of the experiment, that is, in the presence of aldehyde **4** and the ten hydrazides **2a–2j**. To verify that the amplification results were not due to a kinetic selection by means of target-accelerated synthesis, we added SjGST to the pre-equilibrated DCL. The same equilibrium distribution was achieved as is shown in Fig. 5, with hydrazone **5g** strongly amplified, indicating that the amplified components are the result of genuine thermodynamic selection.

Further controls involved a BSA control experiment, which was negative, and DCL synthesis in the presence of a large excess of the nonselective GST inhibitor ethacrynic acid. Component amplification was completely suppressed for both SjGST and hGST P1-1 DCLs, indicating that the GST active site is saturated by the ethacrynic acid and cannot influence the DCL equilibrium composition.

We completed our protein-directed DCL studies by preparing a catalytically inactive SjGST mutant. It was of interest to see whether a functionally disabled enzyme would exert the same control and selectivity on DCL composition as the wild-type enzyme. The conserved Tyr 7 active site residue is known to play a critical role in GSH conjugation for the Sj class of GSTs, stabilizing the GSH thiolate anion through H-bonding from the phenol group, with enzymes lacking this residue being catalytically inactive[52]. We prepared a Y7F mutant of SjGST, in which the crucial tyrosine residue is replaced with phenylalanine. We observed essentially zero activity with this mutant in CDNB conjugation when compared with the wild-type SjGST. However, SjGST Y7F proved equally effective in controlling DCL composition, showing a clear preference for the same thiophene derivative **5g** as was amplified by the wild-type SjGST (see Supplementary Information).

Biological assay was then performed to establish whether the best binding compounds in the GST-directed DCLs were also the best inhibitors of the GST enzyme. To fully explore the isozyme-specific amplification effects of the two DCLs, we separately synthesized hydrazone conjugates **5a–5j** for study. We first confirmed that the amplified ligands **5c** and **5g** bound to SjGST and hGST P1-1 by isothermal calorimetry (ITC) (see Supplementary Information). We then studied their inhibitory activity towards SjGST and hGST P1-1 using the CDNB conjugation assay. The IC_{50} values were slightly higher for all hydrazones against hGST P1-1 compared to SjGST (data ranging from 59 to 126 μM and 22 to 63 μM, respectively; see Supplementary Information). For each isozyme, the DCC amplified hydrazone was the most active; thiophene **5g** had the lowest IC_{50} value (22 μM) among all the library members against SjGST, and *t*-butylphenyl 5c had the lowest value among the four conjugates tested against hGST P1-1 (57 μM). The DCL hydrazone selection process has successfully extended inhibitor structure in the GST H-site, increasing potencies by sixfold for hGST P1-1 (331 to 57 μM) and by over tenfold for SjGST (279 to 22 μM) relative to the starting anchored aldehyde **4**.

Steady-state kinetic studies on the two amplified DCL components **5c** (hGST P1-1) and **5g** (SjGST) confirmed the expected competitive inhibition profile, with both compounds binding to the GST active sites. It was interesting to note slightly higher K_i values for both compounds when assayed against CDNB, a substrate for the H-site of the enzyme, relative to the endogenous G-site ligand GSH. The affinity of the two hydrazone conjugates towards both GST G-sites was relatively close (data ranging from 5.25 to 7.19 μM), as would be expected for two compounds sharing a common GSH-tagged nitrobenzene fragment.

To obtain some molecular insight into the selectivity of our isozymes towards the two hydrazone inhibitors **5c** and **g**, we carried out a molecular modelling study. We surveyed the available GST structures in the protein data bank (PDB) and retrieved those that contained a bound GSH-based ligand. The binding sites of these structures, together with the bound ligands, were aligned, and it became evident that the glutathione portions overlaid well, being bound in very similar conformations in the G-sites (Fig. 6a). In contrast, the conjugate parts of the various ligands showed great diversity in their conformations within the H-site, an unsurprising result given the respective functions of the G- and H-sites. Detailed analysis of the superimposed crystal structures identified the GSH conjugate of 1,2-epoxy-3-(*p*-nitrophenoxy)propane (EPNP) (**6**) bound to cGST M1-1 (PDB code 1c72)[53] as the ligand

that projected functionality into the H-site with the most similar geometry to the energy-minimized structure of hydrazone **5g** (Fig. 6b).

Analysis of the GST–EPNP complex shows the EPNP moiety orienting towards R107 and Q165 in the H-site of the enzyme (Fig. 7c). The side chains of R107, F110, Q165, Q166 and F208 define the pocket that confines the EPNP moiety. On this basis, we could generate a binding model for SjGST with thiophene hydrazone **5g** and for hGST P1-1 with *t*-butyl hydrazone **5c** (Fig. 7). The interactions in the generated binding modes for SjGST in complex with **5g** (Fig. 7a) and for hGST P1-1 in complex with **5c** (Fig. 7b) between the glutathione moiety and the proteins are identical to those reported in previous publications[54,55]. We predict that the hydrazone group of **5g** forms hydrogen bonds to R103 and Q204 in SjGST, and equivalent interactions are observed for **5c** in complex with hGST P1-1. Residue V161 in SjGST and I161 in hGST P1-1 make hydrophobic interactions in our models with the ligands **5c** and **5g**.

As expected, the sub-pockets of both isoforms accommodating the hydrazones are rather hydrophobic, and complement the hydrophobic hydrazones amplified from the DCL. The thiophene hydrazone fits easily in the SjGST binding pocket, with only minor sidechain adjustments necessary (root mean square deviation (RMSD) 0.3 Å between model and crystal structure template), whereas the *t*-butylphenyl group would lead to a steric clash and would require some degree of induced fit in order to bind. Induced fit is also required to accommodate this ligand in the hGST P1-1 pocket, but in that case the binding mode could be stabilized by additional lipophilic interactions of the *t*-butyl group with Y103, H162 and I161. It is worth noting that Chern and colleagues have reported that mutations in that region of the H-site had a great impact on EPNP binding as a substrate, with mutation of cGST M1-1 Q165 to leucine (V161 in SjGST and I161 in hGST P1-1) reducing k_{cat}^{EPNP} by 59%, although K_m showed only small changes[53]. Because the amino acids in the equivalent pocket of SjGST and hGST P1-1 are not highly conserved, these residues have such a great influence on ligand binding that it is likely that these amino-acid exchanges across the isoforms are critical in determining ligand selectivity.

Conclusions

We have demonstrated that reversible synthesis of acylhydrazones can be compatible with protein targets by using aniline as a nucleophilic catalyst. The many advantages of this DCC tool (ready availability of easily customized building blocks, good kinetic and thermodynamic properties leading to ease of analysis, good biological compatibility forming amide-like linkages) may now be realized with biological targets. Most importantly, the acylhydrazone DCLs are truly adaptive, allowing amplification effects to be simply and directly related to structures present at equilibrium.

The GST enzyme proved extremely effective as a DCL template, with two isozymes from the GST family smoothly integrating with the small molecule assemblies and strongly amplifying the best binding components. The selected hydrazones showed increased inhibitory activity of over one order of magnitude from the starting GSH-tagged benzaldehyde **4**, validating the approach in the context of protein–ligand discovery. Interestingly, a single, small DCL composed of only ten members displays isozyme selectivity according to which variant of the GST enzyme is used as the template.

The study at hand has been deliberately confined to a small number of DCL components so as to thoroughly characterize equilibrium distributions and quantify amplifications with the aniline-catalysed hydrazone method. In principle, much larger hydrazone DCLs may be accessed to thoroughly explore chemical space, both within the GST H-site and for other

biological targets[9,15]. It may not be possible, or even desirable, to accurately characterize the equilibrium distribution of such complex DCLs, but this will not be necessary if one simply seeks to identify prominently amplified components from a ligand discovery perspective.

To gain insight into isoform selectivity, we found that each amplified molecule could be effectively docked into its respective GST H-site, although the fine structural features of the SjGST versus hGST P1-1 H-site that discriminate between thiophene hydrazone **5g** and *t*-butylphenyl hydrazone **5c** are unclear at the present time. Structural determination of the complexes of various GST:GS-hydrazone conjugates will be needed for a deeper understanding of the factors that control H-site selectivity. Work in this area, together with applications of acylhydrazone DCC to other biological targets, is the subject of our current research.

Methods

Aniline catalysis of reversible hydrazone formation

The ten hydrazides **2a–j** (10 × 5 μl, 10 mM, DMSO), aldehyde 1 (2 μl, 10 mM, DMSO) and aniline (10 μl, 1 M, DMSO) were added to a mixture of DMSO (93 μl) and ammonium acetate buffer (845 μl, 50 mM, pH 6.2). The DCL was allowed to stand at room temperature with occasional shaking, and was monitored periodically by HPLC to establish the blank composition until the relative populations of the hydrazones became constant. The pH of all samples was raised to 8 by the addition of NaOH (15 μl, 1 M, aqueous). LC-MS verified that each of the expected hydrazones was present in the DCL (HPLC conditions: column, Luna 5 μ C18(2), 30 mm × 4.6 mm, and Luna 5 μ C18(2), 50 mm × 4.6 mm, in sequence; flow rate, 1 ml min^{-1}; wavelength, 254 nm; temperature, 23 °C; gradient, H_2O/MeCN (0.01% TFA) from 95% to 80% over 6 min, then to 45% over 30 min, and eventually to 5% over 5 min) (Fig. 2d). The DCL was then re-synthesized in the absence of aniline, and the HPLC traces at different time intervals were compared (Fig. 2c).

Templated DCL aldehyde 1

SjGST (111 μl, 180 μM, in potassium phosphate buffer 0.1 M, pH 6.8), the ten hydrazides **2a–j** (10 × 5 μl, 10 mM, DMSO), aldehyde 1 (2 μl, 10 mM, DMSO) and aniline (10 μl, 1 M, DMSO) were added to a mixture of DMSO (93 μl) and ammonium acetate buffer (734 μl, 50 mM, pH 6.2). The DCL was allowed to stand at room temperature, with occasional shaking, for 12 h. The pH of the sample was raised to 8 by the addition of NaOH (15 μl, 1 M, aqueous), and the protein was removed by ultrafiltration using a 10,000 MWCO filter (Vivaspin). HPLC analysis was performed and the traces were compared with the blank composition (HPLC conditions: column, Luna 5 μ C18(2), 30 mm × 4.6 mm, and Luna 5 μ C18(2), 50 mm × 4.6 mm, in sequence; flow rate, 1 ml min^{-1}; wavelength, 254 nm; temperature, 23 °C; gradient H_2O/MeCN (0.01% TFA) from 95% to 80% over 6 min, then to 45% over 30 min, and eventually to 5% over 5 min).

DCL composition was identical, regardless of whether the SjGST was present from the beginning or added after pre-equilibration, but equilibration took more than 24 h in the latter case.

For the hGST P1-1 templated library, the ten hydrazides **2a–j** (10 × 5 μl, 10 mM, DMSO), aldehyde 1 (2 μl, 10 mM, DMSO), aniline (10 μl, 1 M, DMSO) and hGST P1-1 (100 μl, 200 μM, in potassium phosphate buffer 0.1 M, pH 6.8) were added to a mixture of DMSO (93 μl) and ammonium acetate buffer (734 μl, 50 mM, pH 6.2). After equilibration for 12 h,

the DCL was analysed using HPLC. Control experiments were performed using the same equivalents of BSA in place of GST.

Conjugate DCLs

To establish the blank DCL composition, the ten hydrazides **2a–j** (10 × 5 μl, 10 mM, DMSO), aldehyde **5** (5 μl, 10 mM, aqueous) and aniline (10 μl, 1 M, DMSO) were added to a mixture of DMSO (96 μl) and ammonium acetate buffer (839 μl, 50 mM, pH 6.2). The DCL was allowed to stand at room temperature, with occasional shaking, and was monitored periodically by HPLC to establish the blank composition until the relative populations of the hydrazones became constant. The pH of all the samples was increased to 8 by the addition of NaOH (15 μl, 1 M, aqueous). LC-MS verified that each of the expected hydrazones was present in the DCL (Fig. 5) (HPLC conditions: column, Luna 5 μ C18(2), 50 mM × 4.6 mm, and Luna 5 μ C18(2), 250 mm × 4.6 mm, in sequence; flow rate, 1 ml min^{-1}; wavelength, 254 nm; temperature, 23 °C; gradient H_2O/MeCN (0.01% TFA) from 95% to 5% over 40 min).

For re-synthesizing the DCL in the presence of the protein SjGST (278 μl, 180 μM, in potassium phosphate buffer 0.1 M, pH 6.8), the ten hydrazides **2a–j** (10 × 5 μl, 10 mM, DMSO), aldehyde **5** (5 μl, 10 mM, DMSO) and aniline (10 μl, 1 M, DMSO) were added to a mixture of DMSO (96 μl) and ammonium acetate buffer (561 μl, 50 mM, pH 6.2). The DCL templated by hGST P1-1 was synthesized by adding the ten hydrazides **2a–j** (10 × 5 μl, 10 mM, DMSO), aldehyde **5** (5 μl, 10 mM, DMSO), aniline (10 μl, 1 M, DMSO) and hGST P1-1 (250 μl, 200 μM, in potassium phosphate buffer 0.1 M, pH 6.8) to a mixture of DMSO (96 μl) and ammonium acetate buffer (589 μl, 50 mM, pH 6.2). The DCLs were allowed to stand at room temperature for 12 h, after which the pH was raised to 8 by the addition of NaOH (15 μl, 1 M). The protein was filtered off using a centrifuge filter of MWCO 10,000 followed by analysis of the filtrate by HPLC using conditions similar to those listed above.

Molecular modelling

To establish ligand alignment, the superposition of GST ligands was carried out using Relibase + 3.0.0 (ref. [56]). A search was first performed to find binding sites that share a sequence identity between 40 and 100% with the target GST crystal structure 1m9a. The resulting 38 structures with bound ligand were superimposed by using binding site residues only. Finally, the ligands from the superimposed structures were extracted and visually analysed.

To carry out a binding mode prediction with Moloc[57], the SjGST crystal structure (PDB code 1m9a SjGST - *S*-hexyl-GSH complex) and the hGST P1-1-GSH complex crystal structure (PDB code 6gss)[54] were used as starting conformations for binding mode generation. The glutathione groups of the synthesized ligands were mapped onto the glutathione groups of the ligands bound to the crystal structures. The hydrophobic hydrazone groups of the synthesized ligands were oriented towards the cavity, lying at the end of the *S*-hexyl site, as observed for the EPNP ligand bound to cGSTM1-1 (PDB code 1c72). In the next step, the protein in complex with the modelled ligand was minimized, considering the ligand as fully flexible. For the protein all residues were kept rigid, except for the amino acids that define the pocket at the end of the *S*-hexyl site (R103, V106, V161, V162, Q204 for SjGST and R100, Y103, I161, H162, N204 for hGST P1-1).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to thank EastChem for the award of a studentship to V.T.B. and the Marie Curie Early Stage Training Network (Syn4chembio) and School of Chemistry at Edinburgh for awarding a studentship to A.M.C. R.B. is supported by an EC Seventh Framework PrograμMe (FP7/2007-2013) under grant agreement no. 223461. M.F.G. is an Engineering and Physical Sciences Research Council (EPSRC) Leadership Fellow. The authors thank A. Cooper (University of Glasgow) for ITC measurements and helpful discussions. N. Petitjean is thanked for the synthesis of hydrazone–GSH conjugates.

References

1. Lehn J-M. Dynamic combinatorial chemistry and virtual combinatorial libraries. Chem. Eur. J. 1999; 5:2455–2463.
2. Rowan SJ, Cantrill SJ, Cousins GRL, Sanders JKM, Stoddart JF. Dynamic covalent chemistry. Angew. Chem. Int. Ed. 2002; 41:898–952.
3. Corbett PT, et al. Dynamic combinatorial chemistry. Chem. Rev. 2006; 6:3652–3711. [PubMed: 16967917]
4. Ladame S. Dynamic combinatorial chemistry: on the road to fulfilling the promise. Org. Biomol. Chem. 2008; 6:219–226. [PubMed: 18174988]
5. Erlanson DA, et al. Site-directed ligand discovery. Proc. Natl Acad. Sci. USA. 2000; 97:9367–9372. [PubMed: 10944209]
6. Corbett AR, Cheeseman JD, Kazlauskas RJ, Gleason JL. Pseudodynamic combinatorial libraries: a receptor-assisted approach for drug discovery. Angew. Chem. Int. Ed. 2004; 43:2432–2436.
7. Hochgürtel M, et al. Target-induced formation of neuraminidase inhibitors from *in vitro* virtual combinatorial libraries. Proc. Natl Acad. Sci. USA. 2002; 99:3382–3387. [PubMed: 11891312]
8. Zameo S, Vauzeilles B, Beau JM. Dynamic combinatorial chemistry: lysozyme selects an aromatic motif that mimics a carbohydrate residue. Angew. Chem. Int. Ed. 2005; 44:965–969.
9. McNaughton BR, Gareiss PC, Miller BL. Identification of a selective small-molecule ligand for HIV-1 frameshift-inducing stem-loop RNA from an 11,325 member resin bound dynamic combinatorial library. J. Am. Chem. Soc. 2007; 129:11306–11307. [PubMed: 17722919]
10. Gareiss PC, et al. Dynamic combinatorial selection of molecules capable of inhibiting the (CUG) repeat RNA-MBNL1 interaction *in vitro*: discovery of lead compounds targeting myotonic dystrophy (DM1). J. Am. Chem. Soc. 2008; 130:16254–16261. [PubMed: 18998634]
11. Scott DE, Dawes GJ, Ando M, Abell C, Ciulli A. A fragment-based approach to probing adenosine recognition sites by using dynamic combinatorial chemistry. ChemBioChem. 2009; 10:2772–2779. [PubMed: 19827080]
12. Eliseev AV, Nelen MI. Use of molecular recognition to drive chemical evolution 1. Controlling the composition of an equilibrating mixture of simple arginine receptors. J. Am. Chem. Soc. 1997; 119:1147–1148.
13. Hioki H, Still WC. Chemical evolution: a model system that selects and amplifies a receptor for the tripeptide (D)Pro(L)Val(D)Val. J. Org. Chem. 1998; 63:904–905.
14. Wietor JL, Pantos GD, Sanders JKM. Templated amplification of an unexpected receptor for C-70. Angew. Chem. Int. Ed. 2008; 47:2689–2692.
15. Ludlow RF, Otto S. Two-vial LC-MS identification of ephedrine receptors from a solution phase dynamic combinatorial library of over 9,000 components. J. Am. Chem. Soc. 2008; 130:12218–12219. [PubMed: 18714991]
16. Turega SM, Lorenz C, Sadownik JW, Philp D. Target-driven selection in a dynamic nitrone library. Chem. Commun. 2008:4076–4078.
17. Xu S, Giuseppone N. Self-duplicating amplification in a dynamic combinatorial library. J. Am. Chem. Soc. 2008; 130:1826–1827. [PubMed: 18211071]
18. Sadownik JW, Philp D. A simple synthetic replicator amplifies itself from a dynamic reagent pool. Angew. Chem. Int. Ed. 2008; 47:9965–9970.
19. Nguyen R, Allouche L, Buhler E, Giuseppone N. Dynamic combinatorial evolution within self-replicating supramolecular assemblies. Angew. Chem. Int. Ed. 2009; 48:1093–1096.

20. Chichak KS, et al. Molecular Borromean rings. Science. 2004; 304:1308–1312. [PubMed: 15166376]
21. Lam RTS, et al. Amplification of acetylcholine-binding catenanes from dynamic combinatorial libraries. Science. 2005; 308:667–669. [PubMed: 15761119]
22. Au-Yeung HY, Pantos GD, Sanders JKM. Dynamic combinatorial synthesis of a catenane based on donor-acceptor interactions in water. Proc. Natl Acad. Sci. USA. 2009; 106:10466–10470. [PubMed: 19171892]
23. Tauk L, Schroder AP, Decher G, Giuseppone N. Hierarchical functional gradients of pH-responsive self-assembled monolayers using dynamic covalent chemistry on surfaces. Nature Chem. 2009; 1:649–656. [PubMed: 21378957]
24. Fujii S, Lehn J-M. Structural and functional evolution of a library of constitutional dynamic polymers driven by alkali metal ion recognition. Angew. Chem. Int. Ed. 2009; 48:7635–7638.
25. Kindermann M, Stahl I, Reimold M, Pankau WM, von Kiedrowski G. Systems chemistry: kinetic and computational analysis of a nearly exponential organic replicator. Angew. Chem. Int. Ed. 2005; 44:6750–6755.
26. Ludlow RF, Otto S. Systems chemistry. Chem. Soc. Rev. 2008; 37:101–108. [PubMed: 18197336]
27. Shi B, Stevenson R, Campopiano DJ, Greaney MF. Discovery of glutathione S-transferase inhibitors using dynamic combinatorial chemistry. J. Am. Chem. Soc. 2006; 128:8459–8467. [PubMed: 16802811]
28. Ramstrom O, Lehn J-M. In situ generation and screening of a dynamic combinatorial carbohydrate library against concanavalin A. ChemBioChem. 2000; 1:41–48. [PubMed: 11828397]
29. Otto S, Furlan RLE, Sanders JKM. Dynamic combinatorial libraries of macrocyclic disulfides in water. J. Am. Chem. Soc. 2000; 122:12063–12064.
30. Nicolaou KC, et al. Target-accelerated combinatorial synthesis and discovery of highly potent antibiotics effective against vancomycin-resistant bacteria. Angew. Chem. Int. Ed. 2000; 39:3823–3828.
31. Milanesi L, Hunter CA, Sedelnikova SE, Waltho JP. Amplification of bifunctional ligands for calmodulin from a dynamic combinatorial library. Chem. Eur. J. 2006; 12:1081–1087. [PubMed: 16240315]
32. Shi B, Greaney MF. Reversible Michael addition of thiols as a new tool for dynamic combinatorial chemistry. Chem. Commun. 2005:886–888.
33. Huc I, Lehn J-M. Virtual combinatorial libraries: dynamic generation of molecular and supramolecular diversity by self-assembly. Proc. Natl Acad. Sci. USA. 1997; 94:2106–2110. [PubMed: 9122156]
34. Cousins GRL, Poulsen SA, Sanders JKM. Dynamic combinatorial libraries of pseudo-peptide hydrazone macrocycles. Chem. Commun. 1999:1575–1576.
35. Furlan RLE, Ng YF, Otto S, Sanders JKM. A new cyclic pseudopeptide receptor for Li^+ from a dynamic combinatorial library. J. Am. Chem. Soc. 2001; 123:8876–8877. [PubMed: 11535105]
36. Roberts SL, Furlan RLE, Cousins GRL, Sanders JKM. Simultaneous selection, amplification and isolation of a pseudo-peptide receptor by an immobilised N-methyl ammonium ion template. Chem. Commun. 2002:938–939.
37. Liu JY, West KR, Bondy CR, Sanders JKM. Dynamic combinatorial libraries of hydrazone-linked pseudo-peptides: dependence of diversity on building block structure and chirality. Org. Biomol. Chem. 2007; 5:778–786. [PubMed: 17315064]
38. Bunyapaiboonsri T, et al. Dynamic deconvolution of a pre-equilibrated dynamic combinatorial library of acetylcholinesterase inhibitors. ChemBioChem. 2001; 2:438–444. [PubMed: 11828475]
39. Bunyapaiboonsri T, Ramstrom H, Ramstrom O, Haiech J, Lehn J-M. Generation of bis-cationic heterocyclic inhibitors of *Bacillus subtilis* HPr kinase/phosphatase from a ditopic dynamic combinatorial library. J. Med. Chem. 2003; 46:5803–5811. [PubMed: 14667233]
40. Poulsen SA. Direct screening of a dynamic combinatorial library using mass spectrometry. J. Am. Soc. Mass Spectrom. 2006; 17:1074–1080. [PubMed: 16735129]
41. Cordes EH, Jencks WP. Nucleophilic catalysis of semicarbazone formation by anilines. J. Am. Chem. Soc. 1962; 84:826–831.

42. Dirksen A, Dirksen S, Hackeng TM, Dawson PE. Nucleophilic catalysis of hydrazone formation and transimination: implications for dynamic covalent chemistry. J. Am. Chem. Soc. 2006; 128:15602–15603. [PubMed: 17147365]
43. Dirksen A, Dawson PE. Rapid oxime and hydrazone ligations with aromatic aldehydes for biomolecular labeling. Bioconjugate Chem. 2008; 19:2543–2548.
44. Rodriguez-Docampo Z, Otto S. Orthogonal or simultaneous use of disulfide and hydrazone exchange in dynamic covalent chemistry in aqueous solution. Chem. Commun. 2008:5301–5303.
45. Hayes JD, Flanagan JU, Jowsey IR. Glutathione transferases. Annu. Rev. Pharmacol. Toxicol. 2005; 45:51–88. [PubMed: 15822171]
46. Mahajan S, Atkins WM. The chemistry and biology of inhibitors and pro-drugs targeted to glutathione S-transferases. Cell. Mol. Life Sci. 2005; 62:1221–1233. [PubMed: 15798895]
47. Li W-S, et al. Overcoming the drug resistance in breast cancer cells by rational design of efficient glutathione S-transferase inhibitors. Org. Lett. 2010; 12:20–23. [PubMed: 19961147]
48. Murray CW, Rees DC. The rise of fragment-based drug discovery. Nature Chem. 2009; 1:187–192. [PubMed: 21378847]
49. Jao SC, Chen J, Yang K, Li WS. Design of potent inhibitors for Schistosoma japonica glutathione S-transferase. Bioorg. Med. Chem. 2006; 14:304–318. [PubMed: 16275109]
50. Tew KD. Glutathione-associated enzymes in anticancer drug-resistance. Cancer Res. 1994; 54:4313–4320. [PubMed: 8044778]
51. Lyon RP, Hill JJ, Atkins WM. Novel class of bivalent glutathione S-transferase inhibitors. Biochemistry. 2003; 42:10418–10428. [PubMed: 12950168]
52. Andújar-Sánchez M, et al. Crystallographic and thermodynamic analysis of the binding of S-octylglutathione to the Tyr 7 to Phe mutant of glutathione S-transferase from *Schistosoma japonicum*. Biochemistry. 2005; 44:1174–1183. [PubMed: 15667211]
53. Chern MK, et al. Tyr115, Gln165 and Trp209 contribute to the 1,2-epoxy-3-(p-nitrophenoxy)propane-conjugating activity of glutathione S-transferase cGSTM1-1. J. Mol. Biol. 2000; 300:1257–1269. [PubMed: 10903867]
54. Oakley AJ, et al. The structures of human glutathione transferase P1-1 in complex with glutathione and various inhibitors at high resolution. J. Mol. Biol. 1997; 274:84–100. [PubMed: 9398518]
55. Cardoso RMF, Daniels DS, Bruns CM, Tainer JA. Characterization of the electrophile binding site and substrate binding mode of the 26-kDa glutathione S-transferase from *Schistosoma japonicum*. Proteins. 2003; 51:137–146. [PubMed: 12596270]
56. Bergner A, Gunther J, Hendlich M, Klebe G, Verdonk M. Use of relibase for retrieving complex three-dimensional interaction patterns including crystallographic packing effects. Biopolymers. 2001; 61:99–110. [PubMed: 11987159]
57. Gerber PR, Muller K. MAB, a generally applicable molecular-force field for structure modeling in medicinal chemistry. J. Comput. Aided Mol. Des. 1995; 9:251–268. [PubMed: 7561977]

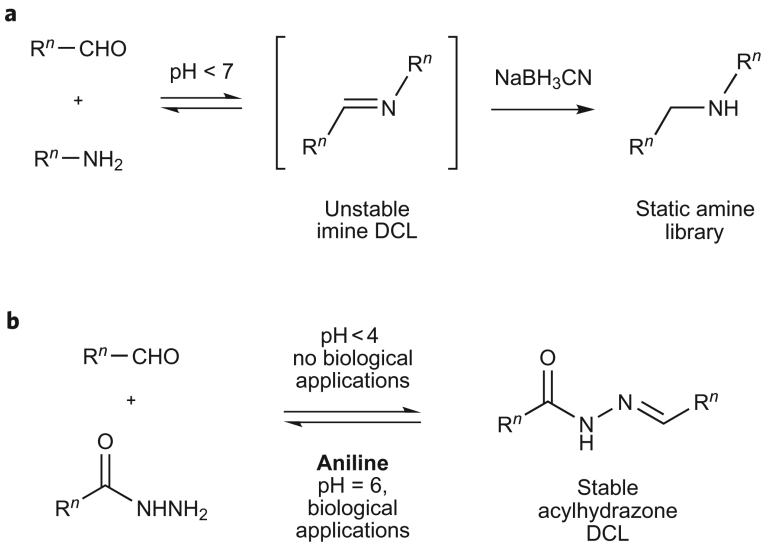

Figure 1. Transimination reactions for DCC

a, Imine DCLs: reversible addition of amines to aldehydes gives unstable imines that cannot be isolated or analysed directly, necessitating an *in situ* reduction step. The resultant static library of amines may or may not share the binding profile of the imine precursors. **b**, Acyl hydrazone DCLs: reaction of aldehydes with hydrazides gives acylhydrazones that have good stability and are amenable to analysis. Equilibration requires acidic conditions that are incompatible with biological targets—a nucleophilic catalyst such as aniline may enable DCL formation at biocompatible pH.

Europe PMC Funders Author Manuscripts

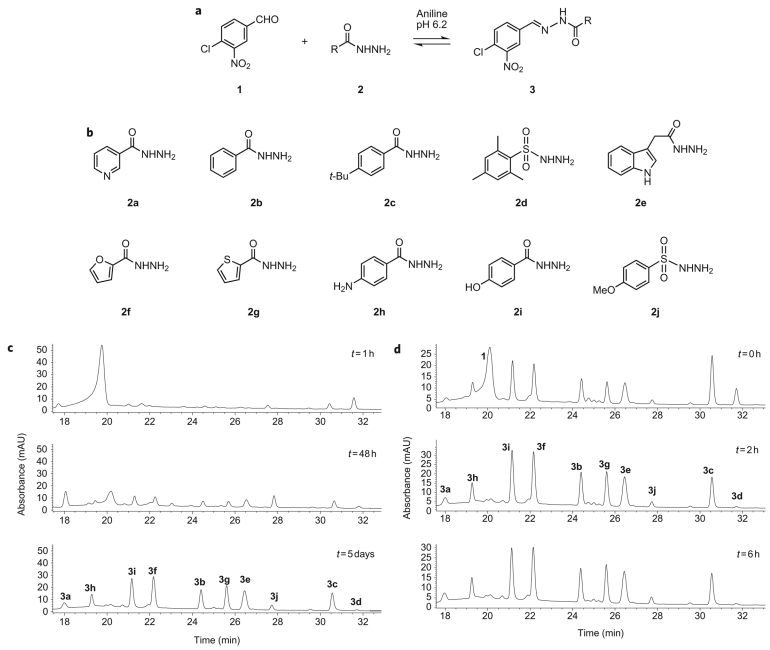

Figure 2. Aniline-catalysed acylhydrazone formation

a, Aldehyde equilibration with hydrazide to form an acylhydrazone. **b**, Hydrazide components of the ten-membered DCL. **c**, DCL established in the absence of aniline. Conditions: aldehyde (5 μM), hydrazides (20 μM each) in NH_4OAc buffer (50 mM, pH 6.2) containing 15% DMSO. **d**, DCL established in the presence of aniline (10 mM).

Europe PMC Funders Author Manuscripts

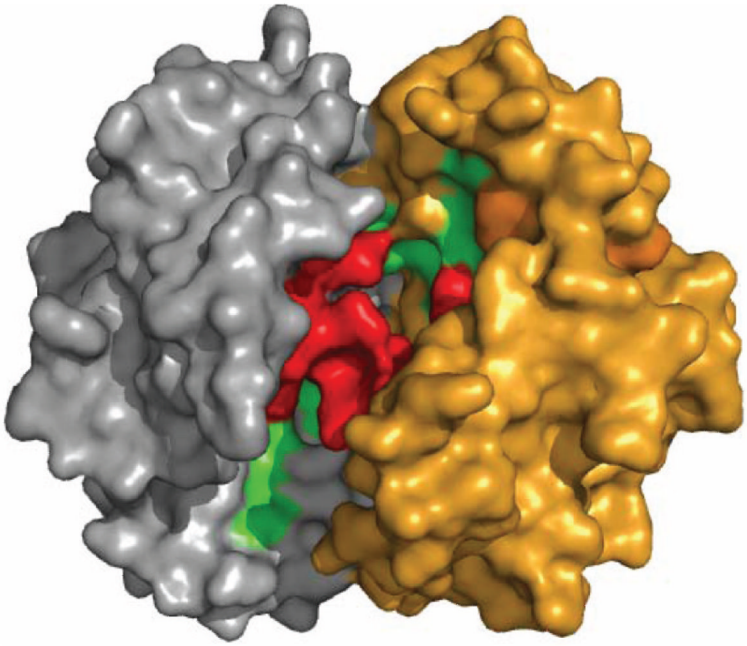

Figure 3. Structure of GST illustrating H- and G-sites
Grey, monomer 1; yellow, monomer 2; green, G-site; red, H-site.

Europe PMC Funders Author Manuscripts

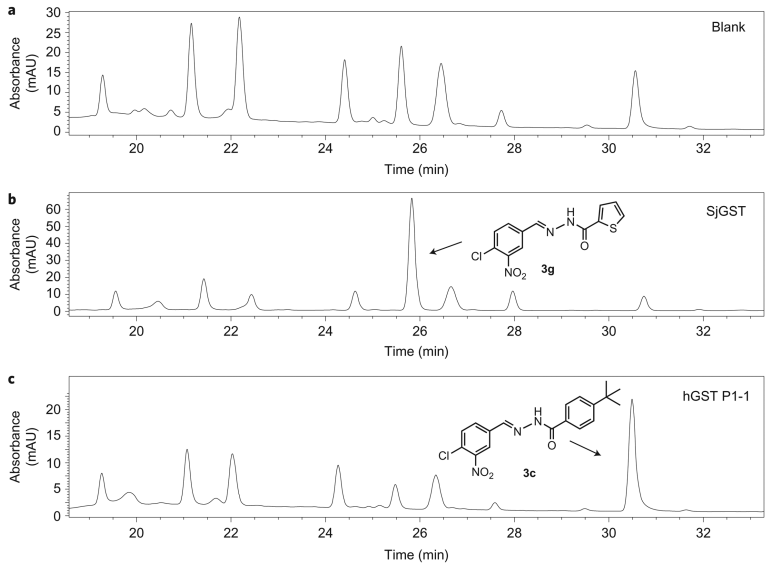

Figure 4. GST-templated DCLs

a, DCL hydrazone composition in the absence of any target (blank). **b**, When the DCL is constituted in the presence of SjGST, the thiophene hydrazone **3g** is clearly amplified. **c**, Changing the target protein to hGSTP1-1 produces a different distrubution, in which the t-butylphenyl derivative **3c** is amplified. Targeted DCL conditions: GST (1 equiv.), aldehyde (5 μM), hydrazides (20 μM) and aniline (10 mM) in NH_4OAc buffer (50 mM, pH = 6.2) containing 15% DMSO for 16 h.

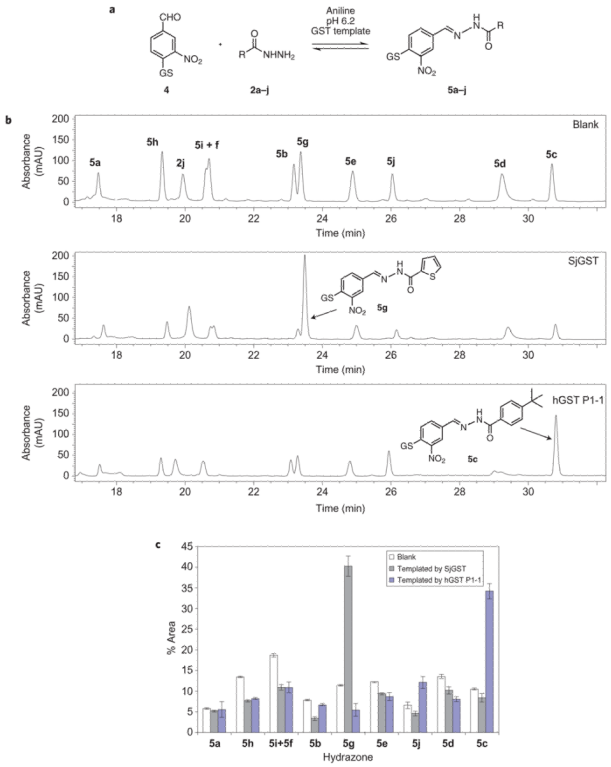

Figure 5. GST-templated DCLs of GSH conjugates

a, Acyl hydrazone DCL based on GSH-conjugated aldehyde **4** (GS S-linked glutathione). **b**, DCL hydrazone composition in the absence of target (blank), in the presence of SjGST and in the presence of hGSTP1-1. DCL conditions: GST (1 equiv.), aldehyde (5 μM), hydrazides (20 μM) and aniline (10 mM) in NH_4OAc buffer (50 mM, pH = 6.2) containing 15% DMSO for 16 h. **c**, Changes in DCL component concentration for blank, SjGST and hGST P1-1 DCLs. The error bars represent the standard deviation over three experiments.

Europe PMC Funders Author Manuscripts

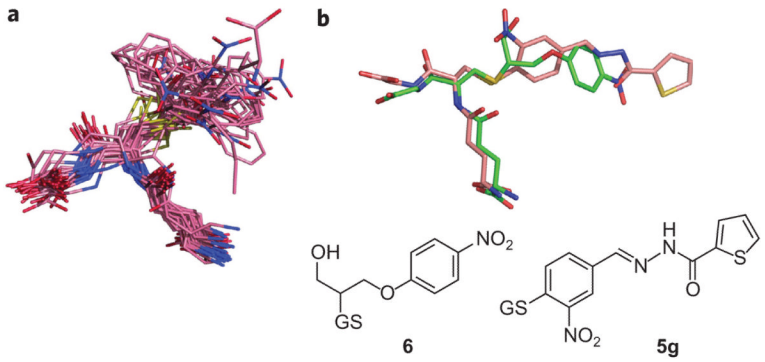

Figure 6. GST ligands

a, Superposition of a selection of GST ligands from the PDB. **b**, Conformation of the GST-bound EPNP ligand **6** as found in the crystal structure of cGST M1-1 (PDB code 1c72, green carbon atoms), relative to the energy-minimized structure of compound **5g** (pink carbon atoms).

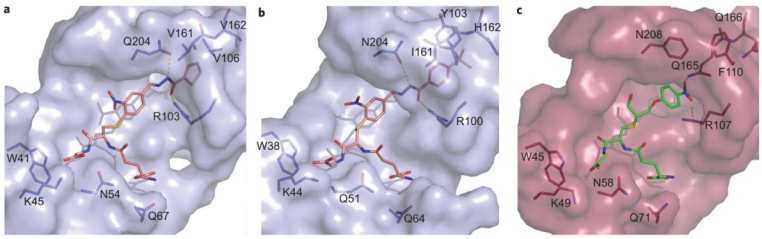

Figure 7. Molecular modelling of amplified DCL components with the GST active site

a, Model of **5g** bound to SjGST. **b**, Model of **5c** bound to hGST P1-1. The binding pocket surfaces are shown in light blue and key amino acids as blue sticks. The ligands are represented in salmon pink, with atoms coloured by type. Hydrogen bonds of the conjugated ligand parts are shown as yellow dotted lines. **c**, The EPNP-cGST M1-1 crystal structure (PDB code 1c72). The binding pocket surface is shown in raspberry pink and key amino acids as red sticks. The ligand is represented in green, with atoms coloured by type. Hydrogen bonds of the conjugated ligand parts are shown as yellow dotted lines.

Europe PMC Funders Author Manuscripts